DOCUMENT RESUME

ED 228 310 TM 830 233

AUTHOR Henk, William A.
TITLE An Analysis of Alternative Cloze Test Formats for Use with Fourth and Sixth Grade Readers.
PUB DATE Apr 83
NOTE 24p.; Paper presented at the Annual Meeting of the American Educational Research Association (67th, Montreal, Quebec, April 11-15, 1983).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Cloze Procedure; Comparative Analysis; Grade 4; Grade 6; Intermediate Grades; Item Analysis; Reading Comprehension; *Reading Tests; *Scoring; *Test Format; *Test Reliability; Test Validity

ABSTRACT

The specific performance characteristics of eight alternative cloze test formats were examined at the fourth and sixth grade levels. At each grade, 64 subjects were randomly assigned to one of four basic treatments (every-fifth/standard, every-fifth/cued, total random/standard, and total random/cued) and tested. Responses on each of the cloze passages were scored in both a verbatim and synonymic manner. Subjects were administered the Ekwall Reading Inventory to determine instructional reading levels. Analysis indicated significant first-order interaction effects for the deletion pattern by scoring mode and blank condition by scoring mode interactions at both grade levels. In the sixth grade analysis, the deletion pattern by blank condition interaction also attained significance. High positive relationships were observed between verbatim and synonymic scoring across all basic cloze forms. Two analyses tested the maximum item independence hypothesis associated with a context distance of four words between blanks. At fourth grade, the results were equivocal; however, at sixth grade, the evidence favored a rejection of this notion. None of the fourth grade cloze formats was significantly related to silent instructional reading levels. At sixth grade, appreciable increases in both the concurrent validity and reliability of the formats were noted. (Author/PN)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED228310

An Analysis of Alternative Cloze Test Formats For Use With Fourth and Sixth Grade Readers

William A. Henk

The University of Georgia

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

W. A. Henk

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TM 830 233

ABSTRACT

In the present study, the specific performance characteristics of eight alternative cloze test formats were examined at the fourth and sixth grade levels. The formats were derived from all possible combinations of the three independent variables, deletion pattern (every-5th versus total random), blank condition (standard versus cued), and scoring mode (verbatim versus synonymic). Separate analyses at each grade assessed (1) the effects of the independent variables on mean cloze response accuracy, (2) the degree of relationship between verbatim and synonymic scoring systems, (3) the effects of varying numbers of words between blank spaces, and (4) the concurrent validity and reliability of the individual cloze formats.

At each grade, 64 subjects were randomly assigned to one of four basic treatments (every-5th/standard, every-5th/cued, total random/standard, and total random/cued) and tested. Responses on each of the cloze passages were scored in both a verbatim and synonymic manner. Lastly, subjects were individually administered silent portions of an informal reading inventory to determine instructional reading levels.

A 2 X 2 X 2 mixed analysis of variance was computed which indicated significant first-order interaction effects for the deletion pattern by scoring mode and blank condition by scoring mode interactions at both grade levels. In the sixth grade analysis, the deletion pattern by blank condition interaction also attained significance.

High positive relationships were observed between verbatim and synonymic scoring across all basic cloze forms. Also, two analyses tested the maximum item independence hypothesis associated with a context distance of four words between blanks. At fourth grade, the results were equivocal, however at sixth grade, the evidence favored a rejection of this notion.

None of the fourth grade cloze formats was significantly related to silent instructional reading levels as determined by the Ekwall Reading Inventory. At sixth grade, appreciable increases in both the concurrent validity and reliability of the formats were noted.

It was concluded that a re-examination of traditional cloze testing appears warranted. In particular, it was recommended that the sixth grade total random/cued/verbatim format receive additional consideration. This determination was based on the format's theoretical capacity to circumvent sampling and scoring biases while maintaining desirable validity and reliability standards.

STATEMENT OF THE PROBLEM

For optimum learning to occur in any reading-related content area, it is essential that teachers provide students with instructional materials that are written at appropriate levels of difficulty. An important step in realizing functional matches between learners and texts, then, is to accurately and reliably assess the reading ability of each individual class member. Through the years, the cloze procedure has come to be regarded as one viable means for doing so. Since cloze tests can be easily and inexpensively constructed, given to whole groups, and scored in a reasonably short amount of time, they are particularly attractive to the already overburdened classroom practitioner. Unfortunately, the model of cloze testing currently in use can be substantively challenged on both theoretical and practical grounds (Bormuth, 1964a; Henk, 1981a, 1982; Huntsberger, 1967; Meredith & Vaughan, 1978; Porter, 1978; Shanahan, Kamil, & Tobin, 1982). For these reasons, the identification and subsequent validation of alternative protocols for cloze test construction and scoring appears warranted.

THEORETICAL AND EMPIRICAL JUSTIFICATION FOR THE STUDY

The overwhelming majority of cloze-related investigations have conformed to the traditional model of (1) deleting every-nth word, where N is a positive integer ranging from 5 to 10, (2) maintaining blanks of an equal length, and (3) accepting only exact word replacements of the original text in scoring. A demonstrated preference for this model persists despite the existence of a considerable body of evidence which militates against its use. There is sufficient reason, for example, to question the almost exclusive reliance on every-nth word deletion patterns. These patterns, which are only approximations of true randomness, have been shown to bias the sample of words that are deleted. Several studies (Bormuth, 1964a, 1964b; Klare, Sinaiko, & Stolurow, 1972; Meredith & Vaughan, 1978; Porter, 1978, Taylor, 1954) have demonstrated the potential for unreliability which exists when approximated schemas are used. In each instance,

significant mean differences were observed across cloze forms constructed from the same passage when the initial deletion point was varied. This lack of equivalence between cloze forms seriously limits the usefulness of the results. If an unrepresentative form is selected from among the total number of possible cloze tests which can be made over a single passage, information regarding students' reading ability will be tenuous at best.

With regard to sampling considerations, the only way to lessen or eliminate the biases introduced by approximated schemas is to institute totally random deletion patterns. In this way, each word in the target passage has an equal and independent probability of being chosen for deletion. This can be accomplished by means of a table of random numbers or some other completely randomized technique. In theory, the results of true random measurement are more objective, and therefore, more generalizable (Huntsberger, 1967). Cloze researchers have been aware of this fact for some time but have been reluctant to implement total random patterns for two basic reasons. The first assumption is that total random test construction is more time-consuming, while the second is that in certain instances the blank spaces might fall too close in proximity to one another to allow sufficient context for processing.

Another highly controversial issue in cloze testing relates to scoring. Typically, only exact word replacements of the original text are credited. While this verbatim scoring virtually assures that different scorers will obtain the same results, the possiblity for rejecting a perfectly acceptable synonymic response is introduced. Considering the diverse number of surface level forms that a single deep structure notion may take, the practice of requiring exact word matches appears highly unrealistic. Even so, most studies to date (Anderson, 1972; Bormuth, 1965; McKenna, 1976; Ruddell, 1964) indicate that no significant differences in rank order result when synonymic scoring is used.

Closely aligned with the issues of both scoring mode and type of deletion pattern is the notion of cueing the response. This cue is provided by the nature of the blank space that is incorporated into the cloze test. The traditional guideline of maintaining blanks of a standard length offers the reader no clue with regard to word length. Since subjects are expected to supply exact word matches, it seems reasonable to provide cues which permit a systematic and objective reduction in the number of competing synonymic alternatives. In many investigations (Anderson, 1971; Emans & Fisher, 1967; Spooncer, 1974) researchers have experimented with different types of cloze cueing strategies. From these studies, only one form, which retained a single dashed space for each letter in the original text word, has shown promise (Entin & Klare, 1979; Henk, 1981b; Rush & Klare, 1978).

PURPOSE OF THE STUDY

While attempts have been made to study the effects of various deletion patterns, blank conditions, and scoring procedures as separate main effects, little is known of the interaction of these factors. Only one study to date (Henk, 1981a) has examined these three factors in combination with one another. In this study, conducted with adult readers, the consolidation of variables into an interactive framework served to highlight desirable cloze format attributes which were inadvertently masked in previous investigations. The present study was an extension of this same basic line of research to fourth and sixth grade readers. It was designed to explore the unique performance characteristics generated by the selected manipulation of deletion pattern, blank condition, and scoring mode factors at these grade levels. Equally important, the concurrent validity and reliability of each alternative cloze format was computed, so that comparisons of functional utility could be made.

DEFINITIONS OF KEY TERMS

Alternative Cloze Test Format--Any of eight possible format combinations derived from each of two types of deletion patterns, blank conditions, and scoring modes used in this study.

Context Distance--The number of words which occur between two adjacent blank spaces in a cloze test.

Cued Blank--The blank condition wherein a single dashed space is retained for each letter in the deleted word.

Equivalence--The state when any of the possible cloze forms which can be generated from a single passage yield similar testing results.

Instructional Level--The level of written discourse at which an individual comprehends 70% or more of an intended message following silent reading.

Maximum Item Independence--When the ability to provide an appropriate response to an individual cloze item does not depend on whether or not a correct answer can be supplied in an adjacent blank space. A context distance of four words between blanks (as in every-fifth word deletion patterns) has been accepted as yielding maximum item independence.

Synonymic Scoring--A system of judging cloze responses based on both grammatical and semantic acceptability with regard to other sentence elements and to the total passage.

RESEARCH QUESTIONS

A total of nine research questions were posed for both the fourth and sixth grade levels. Research questions 1 through 5 addressed the main effects and first- and second-order interactions associated with the three independent variables. Research question 6 dealt with scoring systems, questions 7 and 8

dealt with context distances, and question 9 dealt with concurrent validity. A liberal paraphrasing and summated account of these questions would take the form:

RQ1-5: What are the respective effects of every-fifth versus total random deletion patterns, standard versus cued blanks, and verbatim versus synonymic scoring methods on the cloze test performance of fourth and sixth grade readers?

RQ6: What is the degree and nature of the relationship between cloze tests scored in a verbatim manner and the same tests scored in a synonymic manner?

RQ7&8: Will there be differences in mean response accuracy between context distances of four words and those less than four words (test of maximum item independence hypothesis)?

RQ9: What is the degree and nature of the relationship between cloze test performance and silent instructional reading levels for each of the eight alternative formats?

METHOD

Sixty-four fourth and 64 sixth grade students attending two elementary schools located in a suburban area in southwest Pennsylvania were randomly selected for inclusion in the study. The schools were part of a district-wide system in which the population was predominantly white and of a middle class socioeconomic status.

Two separate experiments were conducted, one at the fourth grade level and the other at the sixth grade level. For each grade, a 2 X 2 X 2 mixed design was used. The two levels of the first factor, deletion pattern, were every-fifth word and total random deletion. The second factor, blank condition, was also comprised of two levels, standard and cued. These first two factors represented

the between-subjects variables. In combination, levels of these between-subjects variables resulted in the following four basic experimental conditions: every-fifth/standard, every-fifth/cued, total random/standard, and total random/cued. Subjects were randomly assigned to treatment conditions (n=16). The third factor, scoring mode, consisted of verbatim and synonymic levels. Since each individual's responses would be evaluated in both a verbatim and synonymic fashion, scoring mode was designated as a within-subjects factor. Its inclusion resulted in a total of eight possible cloze format types.

The reading passages were selected from among those units appearing in the Specific Skills Series (Boning, 1962). All of the basic cloze format types were modified versions of the passages that had been selected to represent the respective grade levels. In the every-fifth word condition, the starting deletion point was randomly determined and then every fifth word in succession was replaced with a blank space. A table of random numbers was used to determine which words would be deleted in the total random condition. In the standard blank condition, a continuous line of 15 PICA typewriter spaces was used to replace the original text word regardless of its length. For the cued condition, a single dashed space was retained for each letter that appeared in the original text word. Examples of the four basic cloze forms which resulted from combining the deletion pattern and blank condition variables appears in the Appendix.

Following testing, the cloze forms were sorted into the four experimental conditions and assessed in two modes. In the verbatim mode, only exact word replacements were credited. For the synonymic condition, each non-verbatim replacement was re-evaluated for synonymic appropriateness using a system consistent with the work of Goodman and Burke (1972) and Vaughan, Tierney, and Alpert (1977).

In the second phase of the experiment, each subject was individually administered the Ekwall Reading Inventory in order to determine independent, instructional, and frustration levels for silent reading comprehension. Instructional levels were converted to grade equivalents to make the data appropriate for concurrent validity analyses.

STATISTICAL ANALYSIS

The main analysis for each grade level addressed Research Questions 1 through 5. To answer these questions, a 2 X 2 X 2 mixed effects analysis of variance procedure was used to determine the impact of the three independent variables on cloze test performance (Kirk, 1968). In this analysis, deletion pattern and blank condition represented between-subjects variables and scoring mode represented a within-subjects variable. Post hoc comparisons for main and interaction effects were made by means of Duncan's New Multiple Range Test. This procedure was used as a follow-up test for each subsequent analysis of variance operation as well. Decisions to reject or fail to reject the null hypotheses associated with the research questions were made at or beyond a .05 level of significance.

Using the data from the main analysis, Spearman Rank-Order correlation coefficients were calculated to assess the relationship between verbatim and synonymic scoring systems for each of the individual cloze test formats, for combined every-fifth and combined total random formats, and for the entire data set. These analyses were an extension of Research Question 6. Next, the effects of varying context distances on cloze performance for each grade level were evaluated in two further analyses, one each for Research Questions 7 and 8. For Research Question 7, a 2 X 2 X 5 mixed analysis of variance was computed in which blank condition was a between-subjects factor and scoring mode and context distance were within-subjects factors (Kirk, 1968). Here the operations were performed

only on discrete context distance data (0, 1, 2, 3, and 4 treated separately) derived from the total random conditions of the main analysis. For Research Question 8, a one-way analysis of variance procedure was used to compare context distances of four as represented by the every-fifth/standard condition with the D 0-3 (collapsed) items as represented by the total random/standard and total random/cued conditions for their respective effects on verbatim cloze performance. Finally, to answer Research Question 9, Pearson-Product-Moment correlations were calculated between scores on each of the eight alternative cloze test formats and subjects' silent reading achievement as measured by the Ekwall Reading Inventory. In addition to these estimates of concurrent validity, reliability coefficients were determined for each cloze format using the Kuder-Richardson Formula 21 (Van Dalen, 1979).

FOURTH GRADE RESULTS

(1) RQ1 - RQ3--Significant main effects were indicated for deletion pattern, F (1, 60) = 63.45, $p < .0001$, blank condition, F (1, 60) = 5.59, $p < .02$, and scoring mode, F (1, 60) = 144.55, $p < .0001$. An inspection of these means indicated that every-fifth word deletions yielded more correct responses than the total random deletions, cueing the blanks produced more correct responses than standard blanks, and more correct responses were made when synonymic scoring was done.

(2) RQ4 & RQ5--Significant first-order interactions occurred for both deletion patterns by scoring mode, F (1, 60) = 6.76, $p < .01$, and blank condition by scoring mode, F (1, 60) = 4.60, $p < .04$. The impact of synonymic scoring was greater in the every-fifth conditions than in the total random conditions. Further, cued blanks had an increased impact on verbatim conditions. No second-order interactions were obtained ($p > .79$).

(3) RQ6--All correlations between verbatim and synonymic scoring systems were significant beyond the .0003 level. Rho values ranged from .78 (total random/cued) to .97 (total random/standard).

(4) RQ7 & RQ8--The maximum item independence hypothesis could neither be accepted nor rejected due to equivocal results between the two context distance analyses.

(5) RQ9--None of the eight alternative cloze test formats achieved significance ($p > .05$) with regard to concurrent validity with silent instructional comprehension levels on an Informal Reading Inventory.

SIXTH GRADE RESULTS

(1) RQ1 -RQ3--While the main effects for deletion pattern and blank condition were not found to be significant ($p > .05$), the main effect for scoring mode did attain significance, F (1, 60) = 330.91, $p < .0001$.

(2) RQ4 & RQ5--All first-order interactions were found to be significant. The values were as follows: deletion pattern by blank condition, F (1, 60) = 4.83, $p < .03$, deletion pattern by scoring mode, F (1, 60) = 4.66, $p < .03$, and blank condition by scoring mode, F (1, 60) = 35.13, $p < .0001$. *Post hoc* analyses indicated that (a) cued blanks had their greatest impact on total random patterns, (b) synonymic scoring affected every-fifth conditions more so than total random conditions, and (c) cued blanks influenced verbatim scoring to a greater extent than synonymic scoring.

(3) RQ6--All correlations between verbatim and synonymic scoring systems were found to be significant beyond the .0001 level. Individual cloze format rho values ranged from .86 (every-fifth/cued) to .96 (total random/cued).

(4) RQ7 & RQ8--Both context distance analyses indicated that the maximum item independence hypothesis (which states that four words are needed between adjacent blank spaces) should be rejected.

(5) RQ9--Six of the cloze formats were found to correlate significantly ($p < .04$) with silent instructional IRI performance. Using the Kuder-Richardson Formula 21, it was determined that five of these six significant formats were also sufficiently reliable ($r_{KR_{21}} > .70$). For each significant format, regression equations were computed for converting sixth grade cloze scores into silent instructional reading levels.

DISCUSSION

With regard to the three main effects, differences were noted between the two grade levels. At fourth grade, both the deletion pattern and blank condition factors achieved significance, while at the sixth grade, both were found to be non-significant. The only similarity occurred for the scoring mode factor as it achieved significance at both grade levels. While the differences in main effect data trends between the two grade levels might be taken as developmental in nature, they must be viewed with caution due to the occurrence of first-order interaction effects.

Since both the deletion pattern by scoring mode and blank condition by scoring mode interactions were significant at both grade levels, the most distinctive interaction occurred at the sixth grade level for deletion pattern by blank condition. This effect was not found to be significant at the fourth grade level. By closely analyzing the means, it was determined that cued blanks resulted in markedly higher performance for the total random conditions. The import of this finding relates to its justifying, in large measure, the decision to combine the deletion pattern and blank condition variables for the purposes of inquiry. Since this same interaction effect was achieved in Henk's (1981a) study with mature readers, the notion that developmental factors may be at work when cloze performance is studied in an integrated manner gains further support.

Closely related to the deletion pattern by blank condition interaction was the test of the maximum item independence hypothesis. According to conventional wisdom, context distances of four words between blank spaces should have generated mean scores that were significantly higher than those produced by distances of 0, 1, 2, and 3 words examined separately and when collapsed. While some support for the hypothesis was found with the traditional model of cloze testing at the fourth grade level, overall the evidence pointed toward a rejection of this notion. At sixth grade, both context distance analyses indicated that the maximum item independence hypothesis was not tenable, particularly when total random deletion patterns were combined with cued blanks in an interactive cloze format. Based on these results, the argument that total random deletion patterns should not be used because they do not allow sufficient context for processing might be dismissed.

Another key finding from the main analysis was that verbatim and synonymic scoring systems were found to be significantly related at both grade levels and for each basic cloze test format. These results are consistent with several studies (Anderson, 1972; Hargis, 1972; Henk, 1981a; McKenna, 1976; Ruddell, 1964) demonstrating that little variation in rank occurs when cloze tests are scored in a synonymic fashion as opposed to exact word criteria. Apparently, synonymic processing is governed by the same fundamental language comprehension strategies necessary to respond to verbatim task demands. For this reason, there appears to be no clearcut advantage in crediting synonyms. In fact, the implementation of synonymic scoring only serves to introduce the possibility that results might vary between scorers.

The final, and probably most important aspect of the study, involved analyzing the individual cloze test formats in terms of their concurrent validity and whole-form reliability characteristics. At fourth grade, none of the eight

alternative formats was found to be significantly related to students' silent instructional reading levels as determined by the Ekwall Reading Inventory (1979). Furthermore, the only form that even approached significance, the total random/cued/verbatim, yielded the lowest reliability estimate. While these data trends suggest that cloze testing is probably unwarranted for identifying silent reading levels at fourth grade, it may very well be that cloze is related to oral instructional levels as indicated by an informal reading inventory (Bormuth, 1968; Kirby, 1967). Recent evidence regarding specific brain functioning (Lassen, Ingvar, & Skinhøj, 1978) suggests that oral and silent reading processes may not be as physiologically similar as once thought; and so, the notion that cloze scores might be more closely related to performance in one mode or the other would appear reasonable. The fact that significance was not achieved for any of the formats may also pertain to the tendency for reading activities conducted prior to fourth grade to be predominantly oral in nature (Strain, 1976; Tinker & McCullough, 1975; Zintz, 1975). It would stand to reason, then, that cloze scores might be somewhat less related to silent reading comprehension at fourth grade than it is at sixth where students have spent considerably more time engaged in silent reading endeavors.

The concurrent validity of the cloze formats was markedly higher for the sixth grade passage. Only the two every-fifth/cued forms (verbatim and synonymic) failed to achieve significance. Table 1 in the Appendix shows the concurrent validity and reliability estimates associated with the remaining six cloze formats. An examination of this table reveals that the cloze format possessing the highest validity coefficient, the total random/standard/verbatim, also possessed a reliability coefficient below the .70 standard of acceptability proposed by Bruning and Kintz (1977). Assuming that the every-fifth/standard forms are susceptible to a sampling bias (Bormuth, 1964a, 1964b; Meredith & Vaughan, 1978;

Porter, 1978; Taylor, 1954), three total random formats remain in "contention" for the designation of most functional cloze test. Of these, only the total random/cued/verbatim format offers a completely objective scoring system, that is, the results will not vary among different scorers. When all relevant test characteristics are considered, this format appears to be the most promising alternative to traditional cloze testing protocols. This conclusion is once more consistent with the Henk (1981a) study using undergraduate subjects.

There would appear to be several advantages to the total random/cued/verbatim format. First, it averts, in theory, the sampling biases that can accompany every-nth word deletion patterns. For this reason, the results should be more generalizable. Further, the interaction which results between verbatim scoring and the presence of cues circumvents the tendency towards depressed mean scores usually associated with the total random/standard forms. As a result, the criticisms of insufficient context distances directed against the total random conditions might be dismissed at the sixth grade level because cued versions produced results that were reasonably equivalent to traditionally accepted cloze test models. In addition, the extremely high positive relationship that existed between scoring systems (.96) indicates that little disruption in ability ranking should occur when verbatim scoring is invoked instead of the psycholinguistically more appealing, synonymic scoring mode.

It might also be argued that the total random/cued forms possess more in the way of content validity than the other alternatives examined in this study. This contention is based on the idea that these formats are more nearly aligned with normal reading. When an individual reads, points of uncertainty simply do not occur regularly at five-word intervals. While these points of uncertainty are not entirely randomly distributed throughout a passage either, it would seem that total random sampling might more readily tap the probabilistic nature of language

as discussed by Pierce (1972) and Rubenstein (1973). Cueing the blank spaces also appears more task realistic in the sense that readers have access to several sources of visual information, one of the most potent being word length (Just & Carpenter, 1980; McConkie & Rayner, 1973). By supplying specific word length cues, the test constructor provides perceptual information that would certainly be available to the reader.

IMPLICATIONS FOR THE CLASSROOM AND RESEARCH

Based on the results of this investigation, it appears that a re-examination of traditional cloze testing methodology is in order. In particular, it is felt that educators and researchers should more carefully evaluate the use of cloze tests as measures of silent reading ability at fourth grade or below the sixth grade level. On a positive note, the total random/cued/verbatim format demonstrated desirable measurement properties at sixth grade. Following a more comprehensive analysis involving an assessment of the format's concurrent validity with other measures such as oral sections of informal reading inventories, equivalent multiple-choice comprehension tests, and standardized reading achievement instruments, it might be responsibly advocated for classroom use. Results of cloze tests using the total random/cued/verbatim format might then be utilized for prediction, grouping, or placement purposes. In addition, future research should examine the performance characteristics of this format across a wide variety of content areas, topics, developmental levels, prose styles, and student ability ranges. Considering the fact that traditional cloze testing methodology has enjoyed 30 years worth of inquiry, it would seem that promising alternatives might warrant extensive examination as well. Viewed in this light, there are several possible research endeavors related to alternative forms of cloze testing which can, and should be undertaken.

SUMMARY

This study provides evidence that the unique combination of deletion pattern, blank condition, and scoring mode variables can result in effects that are masked by main effect studies. Furthermore, it suggests that verbatim and synonymic scoring procedures are closely related in terms of specifying individuals' performance rankings within a group. The study sheds some doubt on the long-standing notions that (1) traditional cloze test construction is both easy and produces reliable results, and (2) four words are needed between blank spaces to ensure maximum item independence. By deleting words completely at random, cueing the blank spaces to specify the exact number of letters in the desired response, and scoring the test in a verbatim manner, reliable results might be obtained that are not necessarily susceptible to item dependence confounding. Most importantly, the experiment demonstrated that selected cloze procedure alternatives can be effective predictors of silent instructional comprehension levels as identified by an informal reading inventory at the sixth grade level. Cloze tests were not found to be particularly valid indicators of silent reading achievement at fourth grade. Since this study was primarily exploratory in nature, a considerable body of additional and iterative research is needed to substantiate all findings and conclusions.

REFERENCES

Anderson, J. Research on comprehension in reading, 8. In D. Bracken & E. Malmquist (Eds.), Improving reading ability around the world. Proceedings of the International Reading Association Third World Congress (Sydney, Australia, 1970), Newark, Delaware: International Reading Association, 1971, 115-121.

Anderson, J. The application of cloze procedures to English learned as a foreign language in Papua and New Guinea. English Language Teaching, 1972, 8, 66-72.

Boning, R. Specific skills series. New York: Barnell-Loft, 1962.

Bormuth, J. Experimental applications of cloze tests. In J. Figurel (Ed.), Improvement of reading through classroom practice, Proceedings of the International Reading Association, Newark, Delaware: International Reading Association, 1964, 9, 303-306. (a)

Bormuth, J. Relationships between selected language variables and comprehension ability and difficulty. (Cooperative Research Project No. 2082), Los Angeles: UCLA, 1964. (b)

Bormuth, J. Validities of grammatical and semantic classifications of cloze test scores. In J. Figurel (Ed.), Reading and inquiry, Proceedings of the International Reading Association, Newark, Delaware: International Reading Association, 1965, 10.

Bormuth, J. Cloze test readability: Criterion reference scores. Journal of Educational Measurement, 1968, 5, 189-196.

Bruning, J., & Kintz, B. Computational handbook of statistics (2nd ed.). Glenview, Illinois: Scott-Foresman, 1977.

Ekwall, E. Ekwall reading inventory. Boston, Mass.: Allyn and Bacon, 1979.

Emans, R., & Fisher, G. Teaching the use of context clues. Elementary English, 1967, 243-246.

Entin, E., & Klare, G. Differential relationships of two versions of cloze tests to vocabulary and reading comprehension. In M. Kamil & A. Moe (Eds.), Reading research: Studies and application. Twenty-eighth Yearbook of the National Reading Conference. Clemson, South Carolina: National Reading Conference, 1979, 68-71.

Goodman, Y., & Burke, C. Reading miscue inventory manual: Procedure for diagnosis and evaluation. New York: Macmillan Co., 1972.

Hargis, C. A comparison of retarded and non-retarded children on the ability to use context in reading. American Journal of Mental Deficiency, 1972, 76, 726-728.

Henk, W. A response to Shanahan, Kamil, and Tobin: The case is not yet clozed. Reading Research Quarterly, 1982, 17(2), 591-595.

Henk, W. Effects of modified deletion strategies and scoring procedures on cloze test performance. Journal of Reading Behavior, 1981, 13, 4. (a)

Henk, W. How important is the shape of a word in specifying its identity? Unpublished manuscript, 1981. (b)

Hunsberger, D. Elements of statistical inference (2nd ed.). Boston: Allyn and Bacon, 1967.

Just, M., & Carpenter, P. A theory of reading: From eye fixations to comprehension. Psychological Review, 1980, 87(4), 329-354.

Kirk, R. Experimental design: Procedures for the behavioral sciences. Belmont, Calif.: Wadsworth, 1968.

Klare, G., Sinaiko, H., & Stolurow, L. The cloze procedure: A convenient readability test for training materials and translations. International Review of Applied Psychology, 1972, 21, 77-106.

Lassen, N., Ingvar, D., & Skinhøj, E. Brain function and blood flow. Scientific American, 1978, 239(4), 62-71.

McConkie, G., & Rayner, K. An on-line computer technique for studying reading: Identifying the perceptual span. In P. Nacke, Diversity in mature reading: Theory and research. Twenty-second Yearbook of the National Reading Conference, 1973, 1, 119, 130.

McKenna, M. Synonymic versus verbatim scoring of the cloze procedure. Journal of Reading, 1976, 20, 141-143.

Meredith, K., & Vaughan, J. Stability of cloze scores across varying deletion patterns. In P. Pearson & J. Hansen (Eds.), Reading: Disciplined inquiry in process and practice. Twenty-seventh Yearbook of the National Reading Conference. Clemson, South Carolina: National Reading Conference, 1978.

Pierce, J. Communication. Scientific American, September 1972, 2-13.

Porter, D. Cloze procedure and equivalence. Language Learning, 1978, 28, 333-341.

Rubenstein, H. Language and probability. In G. Miller (Ed.), Communication, language, and meaning. New York: Basic Books, 1973, 185-195.

Ruddell, R. A study of the cloze comprehension technique in relation to structurally controlled reading material. In J. Figurel (Ed.), Improvement of reading through classroom practice. Conference Proceedings of the International Reading Association, 1964, 298-303.

Rush, R., & Klare, G. Re-opening the cloze blank issue. Journal of Reading Behavior, 1978, 10, 208-210.

Shanahan, T., Kamil, M., & Tobin, A. Cloze as a measure of intersentential comprehension. Reading Research Quarterly, 1982, 17(2), 229-255.

Spooncer, F. Hanging out the cloze line. Reading, United Kingdom Reading Association, 1974, 8, 19-26.

Strain, L. Accountability in reading instruction. Columbus, Ohio: Charles E. Merrill, 1976.

Taylor, W. Application of cloze and entropy measures to the study of contextual constraint in samples of continuous prose. Unpublished doctoral dissertation, University of Illinois, 1954.

Tinker, M., & McCullough, C. Teaching elementary reading (4th ed.). Englewood Cliffs, New Jersey: Prentice-Hall, 1975.

Van Dalen, D. Understanding educational research: An introduction. New York: McGraw-Hill, 1979.

Vaughan, J., Tierney, M., & Alpert, M. A psycholinguistic analysis of cloze responses. In P. Pearson & J. Hansen (Eds.), Twenty-sixth Yearbook of the National Reading Conference. Clemson, South Carolina: National Reading Conference, 1977, 200-202.

Zintz, M. The reading process: The teacher and the learner (2nd ed.). Dubuque, Iowa: William C. Brown, 1975.

Table 1

Combined Concurrent Validity and Reliability Estimates

for Selected Sixth Grade Cloze Formats

Condition	Validity	Reliability
Every-5th/Standard/Verbatim	.50	.76
Every-5th/Standard/Synonymic	.64	.85
Total Random/Standard/Verbatim	.65	.63
Total Random/Standard/Synonymic	.57	.88
Total Random/Cued/Verbatim	.52	.92
Total Random/Cued/Synonymic	.51	.91

Every-5th Word/Standard Blanks

Eskimo dogs _____ fierce, muscular, little animals. _____ big feet are like _____.

Every-5th Word/Cued Blanks

Eskimo dogs _ _ _ fierce, muscular, little animals. _ _ _ _ _ big feet are like _ _ _ _ _ _ _ _ _.

Total Random/Standard Blanks

_____ dogs are fierce, muscular, _____ _____. Their _____ feet are like snowshoes.

Total Random/Cued Blanks

_ _ _ _ _ _ dogs are fierce, muscular, _ _ _ _ _ _ _ _ _ _ _ _ _ _ _. Their _ _ _ feet are like snowshoes.

Answer Key

Eskimo dogs are fierce, muscular, little animals. Their big feet are like snowshoes.

Examples of fourth grade cloze tests.

END

DEPT. OF HEW

NAT'L INSTITUTE OF EDUCATION

ERIC

DATE FILMED

AUG. 12 . 1983